Aalto University
School of Business

Techniques for Mining Transactional Data for Personalized Marketing Actions

Bachelor's Thesis
Pipsa Harno
23.05.2017
Business Technology

Approved in the Department of Information and Service Economy xx.xx.20xx and awarded the grade

Table of Contents

List of Tables and Figures

1 Introduction

Already for some time, it has been stated that targeted marketing is gaining more popularity over mass marketing, especially in the retail sector. Over the years, when technology has become more efficient and powerful, targeting even individual customers with promotions of their like has become possible. At the beginning, segmentation and therefore targeting marketing efforts were only based on geographical clustering, for example, the place of living or working. After that, also demographical characteristics such as age, gender, occupation and income were considered when targeting marketing. (Miguéis, Camanho, & Cunha, 2011).

The benefits of knowing your customers can be divided into three areas: First, getting to know new customers helps in predicting the future since the new customers are the future of the business. Second, how are the products and product categories of a store related to each other in customers purchasing behavior. And third, knowing what kind of customers you have helps in making the experience for each customer individually good and pleasant. (Wei, 2010). For a big retailer, the number of customers and the number of products sold can be huge. Just by using customer relationship management (CRM) all the benefits cannot be gained. This is where data mining comes in by helping to analyse huge amounts of data gathered about the customers. (Wei, 2010).

The objectives of this paper are to describe what kind of data mining techniques there are for the retail sector to use, and how can these technologies be used when planning targeted marketing. The focus of this paper is in the data mining of transactional and demographic data. Some of the recommendation systems for e-shops are also introduced but for example analysing the browsing data of an online shop's web page is left with less attention.

The structure of the rest of the paper is as follows: First, there will be a description of how the data about a retailer's customers is collected and what kind of data is collected and how the situation is in Finnish grocery stores. After that, different kind of data mining tools will be introduced starting from the market basket analysis and following the basics of recommendation systems and different data mining tools used for recommendation systems. Before concluding the thesis, more of a marketing point of view is given about the data mining in general.

2 Data collected by the retailers

The whole process of analyzing transaction data starts from collecting the data. Only a couple of decades ago the development of technology was at a point where it was only possible to store and process reduced datasets with simple models. However, the development of technology has enabled the storage and processing of vast amounts of data nowadays. (Videla-Cavieres & Ríos, 2014).

When planning marketing strategies for targeted marketing retailers often use two types of data. The first type includes demographic data (age, sex, children etc.) and geographic data (place of living, place of working etc.). This kind of data is often collected through questionnaires or if a retail chain has a loyalty card program in use, by asking customers to fill in the information when becoming a member of the loyalty program. The second type of data is the data about the customer's interactions with the retailer. This includes for example the point-of-sales data, web browsing data, answers to surveys conducted and feedback given by the customer. (Bose & Chen, 2009).

Usually, transaction data can be divided into four entities that are stores, customers (if the customer can be identified), transactions (or orders, baskets or item sets) and items (Linoff & Berry, 2011). Most of the time transactional data is in a binary form and therefore it is also the common type of data used in the applications of market basket analysis and other recommendation systems. Binary form suits well since the data is often collected so that the customer either bought the product or not. For market basket analysis, besides only using the data about the products purchased it is also possible to consider contextual information such as time and place to get a deeper understanding of customers' purchasing behavior. (Aguinis, Forcum, & Joo, 2013).

For a retailer, it is also meaningful to think about how detailed the transaction data stored and used should be. For example, merging all the chocolate bars into a category called "chocolate bars" and not specifying the brand, size or flavor of a bar can clarify the market basket analysis. Aggregation can also help finding association rules, reduce the computational effort, runtimes and storage space needed. However, it can vary based on a situation and the goal of an analysis how detailed the data should be. One option is also to first use larger categories, and when interesting association are found to focus on that specific association and use more detailed data. However, it is important to notice that by categorizing some items does not mean that all the items in the analysis must be categorized at the same level. Important, frequent or expensive items should sometimes be kept in a more detailed level. (Linoff & Berry, 2011). By rolling up the items into a higher level helps also when introducing new items. For example, if a retailer starts selling peach juice, it can be added to the recommendations of customers in a way that other fruit juices are also recommended. This method is known as generalizing. (Ahn, 2012).

2.1 The situation in the biggest Finnish grocery store chains

By implementing a loyalty card program, a retailer makes possible to not only follow the point-of-sale data but also link this data to individual customers. Often, when applying for a loyalty card, the customer is asked to fill out some personal information (age, number of children, living place etc.). This information gives a good foundation to categorize customers and find similarities between similar customers. (Linoff & Berry, 2011).

Finnish grocery stores collect a lot of information about their customers, mostly due to their loyalty programs and the fact that customers use their loyalty cards when shopping. The two biggest grocery store chains in Finland are S Group and K-Group with the combined market share being almost 80 per cent. This number is from the year 2016. (Päivittäistavarakauppa ry, 2017). In 2015, slightly over 3,8 million people and 2,2, households in Finland had the loyalty card for K-Group called K-Plussa card, and more than 2 million Finnish people had the S Groups loyalty card, S-Etukortti (Raijas & Järvelä, 2016).

Based on the privacy statement in accordance with sections 10 and 24 of the personal data act (523/1999), all the customer information collected by a company must be clearly stated for the customers. K-Group’s privacy statement includes for example purchasing data collected at total sum of the purchase, product group and product level if the customer has not denied collecting these. Personal data collected includes for example the beginning date of the customer relationship, name, address, phone number, email address, birthdate, gender, language, number of children in the household and their birth years. Information related to the use of the loyalty card include for example date, time and place of all purchases, discounts granted, and of course, all the products/product categories purchased. Also, all the contacts with customer service are recorded and stored. With a special agreement from a customer, K-Group can also collect information about the movements of a customer in certain stores or close to the stores. This can be done with the help of Bluetooth or Wi-Fi connection. (Kesko, Contractual terms and conditions).

For the information collected through web browsing there are some contact and identifiers information and by using cookies or similar technology K-Group can follow which K-Group companies' pages and products a visitor has been browsing. The reasons why K-Group collects all these information about its customers are also mentioned in the privacy statement. K-Group states that the information is gathered to improve customer service, marketing, analyses, statistics and to contact customers. (Kesko, Contractual terms and conditions).

For S Group, the information gathered does not differ much, however, there was no mentioning about gathering location based information about customers. In the privacy statement of S Group It is stated that the information is gathered for the S Group to manage its customer relationships, provide better services and marketing for its customers and to improve the business in general. It is also stated that S Group uses product level data for marketing purposes if the customer has not specifically denied it. (SOK, Asiakasomistaja- ja asiakasrekisteri).

3 Techniques for data mining of transactional data

There are many techniques for mining transactional data. Market basket analysis is a common technique to analyse associations between different products and product categories in general. Recommendation systems on the other hand enables to find products that a certain customer could be interested in. There are also many data mining techniques that are useful when developing a recommendation systems. These include for example clustering, decision trees and regression analysis. All these techniques will be later described in more detail.

3.1 Market Basket Analysis

For a grocery store to benefit from the large amount of data collected through the loyalty programs, a useful tool to deploy is Market basket analysis (MBA), also known as association rule mining. MBA is a data mining technique which is used in many fields all the way from nuclear science to marketing and retail. When used in the retail sector, the goal is to predict and understand the buying behavior and habits of customer. Market basket analysis does not refer to just one single algorithm or application, but more to a set of applications that can be used to analyze the point-of-sales data in order to find relationships between different products and product categories. (Kaur & Kang, 2016). Since the amount of transactional data is often large, market basket analysis enables finding hidden, nonobvious and even surprising associations between products or categories. (Aguinis et al., 2013).

There are many benefits of using market basket analysis. First of all, by knowing which products are often purchased together enables the recognition of driver products or product categories. Driver products are products that increase the sales of some other product. An example of a driver product is a mobile phone. Often by advertising mobile phones the sales of mobile phone cases also increases. By knowing the driver products, a retailer can increase the effectiveness of advertisement and promotions. (Solnet, Boztug, & Dolnicar, 2016). Recognizing the driver products also helps in planning the arrangement of products in a store. Products that can drive the purchase of another product are often placed next to each other. For products that are often purchased together but are not necessarily driver products, for example milk and bread, it can be beneficial to place them on the opposite sides of a store to make sure that a customer must walk through the whole store in order to get those two products. (Aguinis et al., 2013). It needs to be remembered that market basket analysis is not only about what products the customers purchase together but also what they do not purchase. For example, if a customer buys an indoor plant but does not buy a plant nutrient, there would be a way to increase the sales of plant nutrients with some marketing tools or by placing the products differently in a store. (Linoff & Berry, 2011).

3.1.1 Basic Association rule methods

There are three common methods to find out association rules and understand their strength and nature. These methods are support, confidence and lift (Aguinis et al., 2013).

Support is the easiest one to calculate among these three methods and does not require any advanced skills or programs. There are a couple ways to define the support. The first one is just to count how many item sets there are that includes a certain product. Another possibility is to divide this number with the number of all transactions to get a relative occurrence. Lastly, sometimes support is also defined so that it is the number of a certain items/item sets divided by the number of transactions that have enough items for the rule to apply. In Table 1, two transactions support the rule “if bread, then milk”. When divided with all the transactions the support for this rule is 50 percent and when divided with all the transactions that have enough items for the rule to apply (transactions 1, 2 and 3) the support is two out of three. (Linoff & Berry, 2011). A disadvantage of support is that with datasets containing a large amount of data, for example millions of transactions with thousands of different items, the values of support usually stay low. Since the values tend to be low, however they cannot be below zero they are often very close to each other. That of course reduces the effectiveness of support since the idea is to find differences from the dataset. However, by using product categories (for example all the chocolate bars are just referred as chocolate bars), this problem is decreased. (Aguinis et al., 2013).

Table 1: An example of transactional data

Transaction	Items purchased
1	Apple, bread, banana, milk
2	Bread, milk, yogurt, potatoes
3	Apple, butter, potatoes
4	Milk

As for *confidence*, it measures how well the rule can predict the right-hand side of an association rule (rule being “if bread, then milk” the right hand-side is the “milk”). It tells how likely it is that an item is chosen given that another item has already been chosen. Confidence is also relatively easy to calculate. The confidence of the rule “if milk, then bread” is the ratio of the support to the number of transactions with the item in the left-hand side of the rule. In Table 1, with the rule “if milk, then bread” the confidence would be 2/3 so 0,67. All in all, the confidence tells how well a rule predicts what is on the right-hand side. (Linoff & Berry, 2011). An advantage of confidence compared to support is that it can be also used in large and rich data sets since it only considers the transactions that include items of the user’s interest. Another advantage is the causality when counting the confidence twice, first with a rule “if A, then B” and then with a rule “if B, then A”. This can lead to a drastically different values. (Aguinis et al., 2013).

The third common method to measure how good an association rule is *lift* (also called improvement). In general, lift tells if there is an association at all between certain items and whether the association is positive or negative. The formula to calculate lift is as follows: $\frac{P\ (A \cap B)}{P\ (A) * P\ (B)}$.

By using Table 1 and calculating the lift for the rule "if milk, then bread" is $\frac{2/4}{3/4 * 2/4} = 1{,}33$. If the value of lift is below one, there is a negative association between the two items. This means that if there is an item A in a transaction it is likely that there is no item B. This could be the case for example with unhealthy food and some fat-free products. If the value of lift is close to zero, there is no association between the two items and if the value is more than one, the association is positive. With a positive association, the presence of A is associated with a presence of B. With the rule "if milk, then bread" the lift is positive and it can be said that there is a positive relationship between these two items. By calculating lift, it can be said whether it is only a chance that the two items appear together or if there is something more in it. The downside of using lift is however the personal preference of how much above or below one the value should be to have a statistically significant association and therefore which item sets are selected for further research. (Aguinis et al., 2013).

3.1.2 The Apriori algorithm

The Apriori algorithm is one of the most common tool for mining transactional data to find frequent sets of products and association rules. The Apriori algorithm can be used for the data that already satisfies a certain level of support and confidence set by the user. The idea of the Apriori algorithm is relatively simple. First, the algorithm counts the occurrence of different items in the transactional dataset. After finding frequent item sets of one item the algorithm continues to find bigger item sets with the one frequent item found in the first phase. The idea is that in order for the item set "apples and milk" to be a frequently occurred, both the apples and milk have to be frequently occurred on their own as well. This kind of iterative approach is also known as level-wise search. (Emrah, Kantardzic, & Cakir, 2015).

A common application of the Apriori algorithm is a pruning method called minimum support pruning. In general, pruning is a technique that makes finding association rules easier. It starts with the whole transactional data but reduces items and item sets step-by-step when the item sets do not meet a certain criterion. In the minimum support pruning, the criterion is a level of support for a rule. For example, having a million transactions and a minimum level of support 1 percent, all the rules that are not supported by 10 000 transactions are reduces during the first step. Having a rule “if A and B, then C” would mean that by using the minimum support of 1 percent all A, B and C must appear in at least 10 000 transaction for the rule to apply. Also, all the combinations of items A, B and C must appear in at least 10 000 item sets for the rule to be true. (Linoff & Berry, 2011).

3.1.3 Other application of market basket analysis

As mentioned before, the term market basket analysis refers to a lot of methods analyzing the buying behavior of customers. Besides the three most common association rule methods and the Apriori algorithm there are also lot of other applications of market basket analysis. For example, Kaur and Kang (2016) have proposed a new association rule mining algorithm that considers also the changes in the data over time. This kind of periodic mining is a fairly new approach and brings additional value to data mining. (Kaur & Kang, 2016). Kumar and Rao (2006) on the other hand have researched how market basket analysis can be used for planning pricing strategies of supermarkets. For example, how small and large baskets differ and how this knowledge can be utilized. (Nanda & Ram, 2006).

Often the methods of market basket analysis assume that all the products are sold in all the stores of a chain and all the time. Chen, Tang, Shen and Hu (2005) have developed an algorithm similar to the Apriori algorithm, called store-chain association rules, that also considers the location of a store and the time when a certain product has been bought. This is a useful method when analyzing multiple store environments that assumes that not all the stores have exactly the same items on sale, and that some of the products might be just seasonally sold. Because of geographical, environmental or political reasons, some products may not be sold in some stores or the seasons they are sold might differ because of those reasons. For example, the period of selling snow pushers can vary a lot between the north and south of Finland. By considering the different subsets of stores and the time periods when products are sold can be useful when analyzing the transactional data to make marketing strategies or inventory strategies. (Chen, Tang, Shen, & Hu, 2005).

3.2 Recommendation systems

Besides market basket analysis, there are also other approaches, for example different kinds of recommender systems that can be used to find products that the customer might be interested to purchase. When the market basket analysis can easily be used for both, the traditional brick-and-mortar stores and online stores, recommendation systems are more often used for online stores. One of the biggest difference between mining the data from physical store compared to an online store is that while physical stores mostly utilize customers' personal information and purchasing history information, online stores can also utilize the browsing history of a customer in addition to the personal and purchasing history information (Mitra, Ghosh, Basuchowdhuri, Shekhawat, & Saha, 2016).

The process of providing recommendations can be divided into three basic stages, first one being understanding the customers, second, delivering the recommendations, and third, measuring the effectiveness of the recommendations. Based on these three stages, the technical implementation of recommendation systems can be divided into six stages. The first one is the collecting of data. The data can be collected from various channels, for example phone, web, questionnaires etc. The methods used to collect information about customers can be divided into two categories: explicit and implicit. Explicit information collecting methods are those that directly ask from the customer for certain information, whereas implicit methods collect the information by observing customers' behaviour. For the implicit information collecting methods, there are three common measures that are often used: The click stream data where clicking a product is seen as being interested in the product. The second measure is the time spent on the page of a certain product, and the third on is the information about the customer's friends assuming that they have similar preferences or a possibility to influence the customer. Most of the time the best recommendations are generated when both the implicit and explicit methods are used together. (Adomavicius & Tuzhilin, 2005).

The next two stages of the technical implementation of a recommendation system are to utilize the data collected to create a customer profile, and to use that data to match products and the profiles in a way that generates the best recommendations. Following the profile creation and finding the right matches for customers is the fourth stage, where the recommendations generated are delivered to the customer. Once the recommendation is delivered, it is important to evaluate the efficiency of the recommendations and in the final stage, use the result of the evaluation to iteratively improve all the previous stages of the implementation process. This kind of a virtuous cycle is good for two reasons. First, it improves the recommendation system cycle after cycle and secondly, it is important to keep up with the changing taste and preferences of customers. (Adomavicius & Tuzhilin, 2005).

One of the biggest advantages of online stores compared to brick-and-mortar stores is that in the online environment, the whole store can be personalized for an individual customer. In a physical store, the amount of shelf space is limited and can only be filled with the most popular items among all the customers. However, in an online store, it is easier to have a wider collection of items in sales and with a recommendation systems or other data mining techniques, the items that first appears to a customer can be items that the customer is likely to purchase. On the other hand, since the number of items in an online store can be so vast, it is almost inevitable to recommend items. (Leskovec, Rajaraman, & Ullman, 2014)

The core technologies recommendation systems use to provide recommendations can be divided into two categories: content-based systems and collaborative filtering systems. The content-based systems recommend products based on the characteristics of a product. For example, if a customer has been purchasing a lot of lactose-free products the recommendation algorithm would provide recommendations about other lactose-free products for the customer. The collaborative filtering systems are founded on the relationships of customers and items. This kind of a system recommends items that are often purchased by similar customers. (Leskovec et al., 2014).

3.2.1 Content-Based Recommendations

For the content-based recommendation systems, first the retailer should define the set of characteristics an item has, that is the profile of an item. For a grocery store, these profiles could include characteristics like cheap/expensive, healthy, organic, lactose-free, the package size etc. Often a vector approach is used for content-based recommendations that calculates the difference of two items based on the differences and similarities of their characteristics. (Leskovec et al., 2014). If an item has a certain character it is marked as 1 and if not with 0. For example, a certain yogurt can be fat-free, not organic and lactose-free. Then the values for these characteristics would be 1, 0, 1. With numerical values, it is possible to either use the exact value of a feature or to consider values similar if they do not differ a lot. When the values have been collected, each item is treated as a vector and with that the cosine is calculated. The cosine tells about the similarities of these items. (LVN, Wang, & Raj, 2014)

After forming the vectors for items there must be also vectors for the customers to be able to calculate the cosine of the user vector and an item vector and hence seeing how big is the difference between the preferences of the customer to the item. The vector and therefore the profile of a customer can be created as follows: if a customer has bought a certain yogurt the value for the yogurt is 1. If 30% of the yogurts the customer has bought have been strawberry flavoured the component of strawberry is 0,3 in the customer profile. (Leskovec et al., 2014).

If the values are numerical, for example the size of a package for meat products, then the user profile can be calculated with normalized values so that the size of a package is subtracted with the average package size of all the meat packages the customer has purchased over time. For example, if the average weight of a package of meat a customer has purchased is 400g and among all those meat packages three is chicken with the package sizes of 250g, 500g and 700g. For counting the component for chicken will be done by subtracting the average for all the meat products from each chicken product: 250g – 400g, 500g – 400g and 700g – 400g. In order to get the value for the component, the average from the subtracted values is counted. For this example, the value would be 83,33 grams. By forming vectors based on the customer profile and an item profile, these vectors can be compared by calculating the cosine and hence telling if the customer would like the item. (Leskovec et al., 2014).

There are several shortcomings when using content-based recommendations. The first shortcoming is that the analysis is often relatively shallow when only the information from previous purchases are used and the customer profile is created only based on that. Another shortcoming is over-specification, meaning that since only the items similar to previous purchases are being recommended the scope of recommended items remains quite narrow. The third shortcoming is eliciting feedback meaning that the quality of recommendations can only be improved when the customer purchases, rates or views a new product. (Balanovic, Marko; Shoham, Balabanović, & Shoham, 1997).

3.2.2 Collaborative Filtering

In a nutshell, in collaborative filtering the process of recommending starts by profiling the customers, then by finding customers with similar profiles and lastly, recommending products to a customer that similar customer have also liked. (Leskovec et al., 2014).

A simple collaborative filtering technique is based on the Jaccard Distance. It is a good measure for the transactional data where we only have the information whether the customer has bought the item or not. An example of the Jaccard distance is as follows: As seen in the Table 2, customer A has bought five items, bread, butter, milk, bananas and soda and customer B has bought seven items, bread, bananas, pasta, ham, orange juice, yogurt and tomatoes.

Table 2: Utility matrix for Jaccard Distance calculations

A	bread	butter	milk	bananas	soda	-	-
B	bread	bananas	pasta	ham	juice	yogurt	tomatoes

The clause for the Jaccard distance is then as follows: $J(A,B) = 1 - (|A \cap B|/|A \cup B|)$. So, for the example above, the Jaccard distance would be $J(A,B) = 1 - (2/10) = 0{,}8$. The higher the result the further apart the further apart the customers are. The drawback with the Jaccard distance is however that is does not consider whether the customer liked the item or not, but only whether the item was purchased or not. (Leskovec et al., 2014).

For example, in online stores, it is often possible to rate the products purchased. This way, recommendations can be provided by the same way than explained before in content-based recommendations. The only difference is that the vectors are made for the customers based on the ratings they have given to different products and customer profiles are compared to each other with the cosine clause. The larger (positive) the result is, the smaller the angle between the two vectors is and therefore, the more similar the two customers are. (Leskovec et al., 2014).

There are two types of collaborative filtering algorithms: memory-based and model-based algorithms. Memory-based algorithms use the entire database to create recommendations. Each customer is compared to other customers in order to find the most similar customers and generate the recommendations through those customers' purchases. Neighbour-based algorithms are examples of these. However, generating recommendations by using the entire database can be slow especially if the database is large. In these kinds of situations, model-based algorithms can be more effective to use. As the name model-based already suggests, the algorithms work so that based on the dataset the algorithm creates a model that can be used to generate recommendations. This way, it is not necessary to use the entire database every time. Examples of model-based algorithm are Bayesian clustering and classification-based algorithms. (Vucetic & Obradovic, 2005).

There is a trade-off whether we compare the profiles of a customer to other customers' profiles or whether we compare items that a customer has bought to other items. When comparing one customer to another customer the results give all the items that should be recommended at once. If a customer A has purchased items A, B, C, D, E, F and G and customer B, whose profile is similar to the profile of customer A but has not purchased products D and E, customer B can be recommended with those products. However, when we are comparing items to other items, each item available must be compared to other items separately. Even though comparing items to items means more work it is often more reliable. It is often better to find items that have the same characteristics and therefore are suitable for recommend to a customer than to find customers that are similar for example based on the purchasing history. Even if two customer profiles would be similar to each other there is a big chance that the two customers differ what comes to the preferences of the products not purchased by the customers. (Leskovec et al., 2014).

Like content-based recommendation systems also collaborative-filtering has its shortcomings even though it can be used to solve most of the shortcomings of content-based recommendation systems. As mentioned in the previous section, collaborative filtering does not work properly with people of unusual tastes. Another shortcoming is related to recommending totally new items of a store. New items will not be recommended to customers until many other customers have purchased or rated the item. Before that, the recommendation systems cannot link the item with any similar items. (Balanovic et al., 1997).

3.2.3 Social recommendation systems and hybrid approaches

A third, and relatively new approach besides the content-based recommendations and collaborative filtering is a social recommendation systems. They use social network data to build more accurate customer profiles (Arazy & Kumar, 2010). For example, Amazon has introduced an application that generates recommendations based on the customer's Facebook friends' purchasing history (Li & Karahanna, 2015).

The hybrid approaches are ways of combining the collaborative filtering, content-based recommendations and social recommendations. One example of a hybrid approach is to use both, content-based recommendation and collaborative filtering separately and then combine the results for final recommendations. Another tactic is to use both approaches in a single model, not separately, and this way generate the recommendations. (Adomavicius & Tuzhilin, 2005).

3.3 Other applications of recommendation systems

Besides the core recommendation systems introduced in the previous section, there are set of other data mining techniques that have been used to develop the recommendation systems in general. Park, Kim, Choi and Kim (2012) have classified the techniques into eight categories as follows: association rule, clustering, decision tree, k-Nearest neighbour, neural network, link analysis, regression and other heuristic methods. From these categories, association rule has already been explained in the market basket analysis section and the rest will be briefly explained next.

3.3.1 Clustering

As the name of the method "clustering" already suggests, clustering is a data mining technique used to group similar data objects into smaller groups, called clusters. As a data mining technique, clustering refers to a large number of methods, for example k-means and self-organizing map, which are some of the most popular clustering methods (Park, Kim, Choi, & Kim, 2012). However, most of the clustering algorithms can be divided into two categories; partitional or hierarchical.

Partitional clustering means that each object can belong to only one cluster, so that the clusters do not overlap with each other. Hierarchical algorithms on the other hand divide data objects into nested clusters. Hierarchical algorithms can be further divided into two categories: agglomerative or divisive. In agglomerative algorithms, all the objects have their “own clusters” at the beginning from where they are merged into bigger clusters until all the objects are in one big cluster. For the divisive clustering algorithms, the idea is very similar, however, now the objects are first in one big cluster, which is then iteratively divided into smaller clusters in a way that each cluster is always split into two new clusters. (Miguéis et al., 2011)

There are also other ways to categorise different clustering methods. One approach is hard vs. fuzzy algorithms where hard clustering algorithms divide the data into distinct clusters so that each object can only belong to one cluster. Fuzzy algorithms on the other hand, can allocate objects into several clusters and by showing the degree of membership, the algorithm indicates how well the object belongs to the cluster. Yet other ways to classify the clustering algorithms are deterministic vs. stochastic (whether the objects always arrive at the same clustering or it uses stochastic elements to find a good clustering), monothetic vs. polythetic (single- or multiple-feature-based ways to divide objects into clusters) and incremental vs. nonincremental (whether all the objects can be clustered at the same time or they will be clustered as they arrive). (Jain, Murty, & Flynn, 1999).

When using clustering in the retail sector, it can be used to either cluster customers into meaningful groups or to find groups of items with similar profiles. When clustering is used in marketing or customer relationship management it is often referred as “segmentation”. When customers are segmented for retailers’ purposes they are often segmented based on geography, demographics, buying behaviours, revenue or based on the customer’s answer on different kinds of surveys conducted. For clustering items, the market basket analysis is often used but hierarchical clustering is a useful tool as well. With clustering, a retailer can cluster groups of items that usually occur together in customers’ purchases. (Linoff & Berry, 2011).

3.3.2 Decision trees

Decision trees are a technique for classification. With the help of different clustering methods customers can be grouped based on their purchasing behaviours. When the customers are grouped, the clusters can be profiled by their characteristics with the help of decision trees. After classifying the groups with a decision tree, association rules can be used to find the frequent products purchased by the customers from each group. By going through all these steps, marketing can be targeted for the customer based on his/hers previous purchases but also based on the information of the customer segment's purchasing habits. (Miguéis et al., 2011).

For retailers, a decision tree can be a useful tool when classifying market segments and identifying the features that differentiate one cluster from another. A decision tree is built with nodes (leafs and decision nodes) and branches. Each leaf is a final cluster whereas in a node it must be decided which branch to choose to continue to a leaf (or another node). Where to continue is chosen based on the values of nodes and leaves and which one of these values corresponds the best with the item of interests. The use of a decision tree is started from the root and by following the right branches it leads to a leaf with the correct cluster and its name. (Miguéis et al., 2011).

3.3.3 k-Nearest neighbour

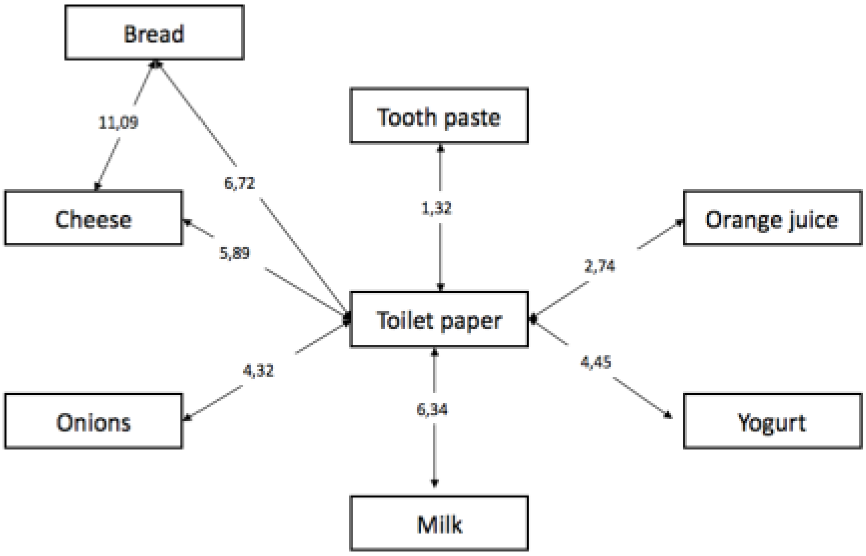
The k-Nearest neighbour (k-NN) model is the most traditional collaborative filtering based recommendation system. The process of making recommendations can be divided into three steps. The first step is to build a customer profile based on either the customer's ratings of products or building the profile based on the purchase information. In the second step, the system finds *k* customers (also called neighbours or recommenders), with similar purchasing or ratings done in the past. Based on the comparison, a neighbourhood of similar customers compared to the target customer is formed so that the more similar the other customer is to the target customer, the closer it is located in the neighbourhood. The third step is to analyse the previously purchased products of the neighbourhood and form a top-*n* item set for the target customer. The top-*n* items are selected so that the purchasing behaviours of closer (similar) neighbours (customers) have more weight than the neighbours further away. These *n* items are the items that the customer should be recommended to purchase. (Kim, Kim, & Ryu, 2009).

3.3.4 Link analysis

Link analysis is often visualized with a graph containing nodes and edges, where the nodes represent objects and edges relationships between these objects. Depending on the field and the goal of the link analysis, nodes can represent for example, people, organizations or objects. If there were two nodes, A and B, an edge would be the pair of nodes, AB, that are connected by a relationship. There are two types of graphs. The first one is fully connected, where all the objects have a relationship between each other. In the other types of graphs all the nodes do not have a relationship with all the other nodes. The graphs can also be categorized into two based on the type of edges they have. In undirected graphs the edges work always two ways. In direct graphs, an edge from A to B is distinguished from the edge from B to A. It is also not required that both edges exist. The advantage of graphs is the possibility for intuitive relationship visualization. (Linoff & Berry, 2011).

A concrete example on how the link analysis and graphs can be utilised in the field of retail is weighted graphs. It can be a useful tool to visualize for example how often certain products appear together in market basket. In the Figure 1, there is an example of how the weighted graph works. Each node is an item and the edge between two items represents the relationship of items being purchased together. The weight tells the percentage of how many times those two items appear in the same market basket. However, the weighted graph does not reveal for example, whether the sales of tooth paste is driven by the sales of toilet paper or vice versa. (Linoff & Berry, 2011).

Figure 1: An example of a weighted graph with the percentage of transactions containing both items of an edge (Linoff & Berry, 2011)



Another common application of link analysis is social network analysis where the aim is to analyse the relationships and interaction between people, and based on those create recommendations. The idea is that often we trust in the recommendations made by friends and family and the idea of social network analysis is to mimic these word-of-mouth recommendations. (Linoff & Berry, 2011).

3.3.5 Neural network

Neural networks (also known as artificial neural networks) can be used for prediction, estimation and classification problems. At the beginning, artificial neural networks were created to understand the biological neural networks with the help of computers. Besides being helpful for biologists, researchers studying artificial intelligence also quickly realised the potential of artificial neural networks as a possibility to endowing computers with learning abilities. Because of this, the artificial neural networks resemble the idea of biological neural networks. (Linoff & Berry, 2011).

The basic unit of artificial neural network is the artificial neuron. The way how the artificial neuron works is called an activation function, which consists of two steps: combination function and transfer function. Combination function combines the inputs that the artificial neuron takes and produces a single value by combining the inputs. Usually, the inputs are assigned with certain weights and therefore the combination function is the weighted sum of the inputs. This is also the moment where the learning ability of a computer can be utilized by assigning the best values to the weights. After combining the values to a function, transfer function produces the output. There are several transfer functions to choose from. The first one is the step function, which mimics the best the biological neuron. The idea behind it is simple: when the weighted sum is above a certain threshold the function gets a value of 1 and otherwise a value of 0. Other common transfer function types are linear, logistic and hyperbolic tangent functions. (Linoff & Berry, 2011).

Neural networks can be used in the retail sector to predict the probability of a certain purchase. For example, a neural network where there are several possible outputs could be used to predict how likely a customer will purchase from a certain department of a retail store. In this kind of a neural network, the inputs could be for example the last purchase, age, gender and so on. The outputs on the other hand would tell how likely it is for the customer to make his/her next purchase from these departments. By assigning correct weights for the inputs, neural network would generate values for each output (departments). There are several ways how the propensities can be utilized for example in marketing. For example, by marketing products of a department with the highest output can be beneficial. Other options are for example to select the top three departments with the highest outputs, or all departments that have an output above a certain threshold value. (Linoff & Berry, 2011).

3.3.6 Regression model

Regression models are widely used statistical techniques for predictive modelling. They can be used for a number of complex problems but the simplest form of regression models is the best-fit line, where the equation describes the relationship between the input and target variable, both of which have to be numerical values. With the best-fit line it is possible to create a function by using data points that can describe for example the relationship between the length of a customership and the amount of money used during the customership. By generating the linear regression line it is possible to use the input, the length of a customership to explain the target revenue. However, the length of a customership is clearly not the only factor affecting on the revenues. The best-fit line equation can be expanded to a multiple regression, where it is possible to take into account multiple factors. The equation with multiple variables is $\Upsilon = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$, where Y is the target variable and X is the input. Betas are the model coefficients.

Besides the ability to estimate the value of a target variable, a regression model can also be used to analyse the relationships between inputs and the target. The signs of the coefficients (β) tells whether the target value increases or decreases if the corresponding input increases. By looking at the absolute numbers of coefficients, it can be said, which input has the biggest impacts to the target variable (the biggest absolute value of coefficients) and which has the smallest impacts (the smallest absolute value of coefficients). For multiple regression models, it is however very important that the variables are carefully selected since the models do not work well with many inputs. Common sense, domain knowledge, forward selection, stepwise selection and backward elimination are ways of choosing the best input variables.

The regression models described so far, have been linear regression models. As the name already suggests, the equation describes a straight line that has no minimum or maximum values. These linear regression models are suitable for continuous variables that can have unbounded number of values. However, for binary problems, linear regression models cannot be used and therefore there are the logistic regression models. In the logistic regression models, the straight line that is generated in the linear regression models is replaced with a function that can get every value between 0 and 1 but never values smaller than 0 or larger than 1. (Linoff & Berry, 2011).

3.3.7 Other heuristic methods

Heuristic methods refer to methods that have been developed from methods already existing by adding new methods to them. A couple example of these are mixture models and ontology method. (Park et al., 2012)

3.4 Model of recency, frequency and monetary

For marketing purpose, a model called Recency, Frequence and Monetary (RFM) is commonly used. It was introduced by Bult and Wansbeek in 1995. (Miguéis et al., 2011). It is one of the most popular data mining technique used by marketers for segmentation and to find the most valuable customers. (Olson & Chae, 2012).

RFM stands for the words recency, frequency, monetary value and it measures variables of how recent was the last purchase made by the customer, how often the customer purchases during a certain time frame and how much money in total the customer have used during that time frame. A lot of research have been done in this field of marketing and there have been also other variables suggested to implement in the RFM model. For example, customer income or age can be important to include in some cases. (Olson & Chae, 2012).

Decision trees, neural networks and logistics regression however, outperform the RFM models in their prediction accuracy (Olson & Chae, 2012). The benefit of RFM, is that it can be used to analyze the value of a customer and develop marketing strategies based on those values. (Hu, Huang, & Kao, 2013). Often, data mining techniques assume that each customer is equally important and through RFM analysis the resources of marketing can be better allocated to different customers. (Olson & Chae, 2012).

Besides using the actual point-of-sales data about the products a customer will most likely purchase, another way to target marketing is for example rewarding customers that often use a lot of money for a specific retail store or on the other hand, encourage those who only use a small amount of money to increase the money used. This could be done for example by giving discounts like "when you purchase with more than 50€ you get 5€ off". RFM is a helpful tool when conducting marketing strategies like this. (Nanda & Ram, 2006)

4 The process of personalized marketing with the help of data

There is really no question anymore about the importance of targeted marketing over mass marketing. The competition in the retail sector has become more intense over the past decades by the development of online marketing. At least two reasons can be found why the importance of customer profiling has been gaining more and more attention in marketing during the last decades. First, marketing has become customer-centric and to implement this, targeted marketing and services are needed. The second reason is the inefficiency of mass marketing and the high expenses compared to the response rate. (Olson & Chae, 2012).

The process of targeted marketing can be divided into three stages. The first stage is the input, which includes collecting the data that enables to target the marketing actions. The second stage, processing, includes activities such as selecting the target customers and profiling them, what products should be targeted for them and how the marketing should be executed. In this stage, it should also be analyzed how the target customer will generate revenue. There are three ways to estimate how the customer generates revenue and all these should be considered. The first one is the amount of money spent in a response to a marketing action. The second way is by multiplying the response probability with the expected revenue. The third way to estimate the revenue is by using the life time value. This is a good method if there are lot of marketing actions so all those can be considered and compared to the amount of money the customer would use during his/her customership in total. The third stage in the process of targeted marketing is output, and it is about evaluating the performance of marketing actions. (Bose & Chen, 2009).

The methods how to deliver recommendations to customers can be divided into three categories known as pull, push and passive. The pull method is used when the customer must specifically request the recommendations after being notified that the recommendation data is available. In the push method, the recommendations are “pushed” to the customer meaning that the customer does not have to see any effort to get the recommendations. An example of this could be for example sending the recommendations to the customer by email. The third method, passive, is used when the website shows the recommendation on the side of a web page when the customer is browsing on it. Especially the first and third method are more often used in online shops. (Adomavicius & Tuzhilin, 2005).

There are two common ways how to increase the sales by using marketing: cross-selling and up-selling. Cross-selling is a marketing strategy where similar or related products to the ones the customer has already been purchasing are recommended to the customer. For example, “customers who bought this, also bought this” marketing actions are cross-selling. Up-selling, on the other hand, is marketing better versions of products than the customer already has. For example, a customer who has a certain kind of mobile phone can be suggested to buy a better phone by advertising those phones to the customer. (Bose & Chen, 2009). A simple example of how to utilize cross-selling marketing strategy in marketing is knowing that potato crisps and dip are often bought together, and marketing only one of these products can increase the sales for both, crisps and dip. By placing complementary items like these close to each other in a store increases the likelihood of the customer buying from both of the categories. (Aguinis et al., 2013).

A concrete example of a market basket analysis used in targeting market actions is to analyze whether in a multilingual area, a grocery store should advertise its products in all the languages used in the area. For example, a study about Texas revealed that there are differences in the buying habits of English speaking and Spanish speaking population and therefore different products should be advertised in different languages. (Linoff & Berry, 2011)

5 Conclusions and prospects for the future

All in all, there has been relatively much research conducted in the field of data mining and also more specifically in the field of retail. Because of the development of technology, larger datasets about customers' purchasing behaviour can be stored, processed and analysed. The most common tool for analysing the transactional data is market basket analysis, which can be used for both bricks-and-mortar stores and online stores. For the online store environment, there is a lot of different recommendation system applications that can also be used. All of these efforts are often used for planning marketing strategies and to make both the service itself and marketing more targeted to the customers. A big issue with collecting personal data about customers is of course the privacy and the safety of the data. This, of course needs to be carefully addressed when collecting and using data about customers, however, in this paper the area is not covered in detailed.

It is a current issue for bricks-and-mortar stores to develop ways to compete with online stores. Online stores can provide for example convenience and broader selection of goods compared to bricks-and-mortar stores. What also makes bricks-and-mortar stores struggle compared to online shops when it comes to personalizing marketing, is that the bricks-and-mortar retailers only have the point-of-sale data. In an online shop, all the clicks a customer makes can be tracked and even if the customer does not buy anything, the customer's path in the web page can still be utilized in targeted marketing. (Keller, Raffelsieper, & Ag, 2014). However, this can change in the coming years. As mentioned in the section about what data Finnish grocery store chains collect, K Group is able to collect location based data about its customers in some business premises. This kind of customer tracking can be done for example with the help of Wi-Fi or RFID sensors. For example, Amazon has created a bricks-and-mortar store where customers log into to an app when entering a store. With the help of the app and the technology in the store, the store can detect what products a customer grabs from the shelves. (Daniels, 2016). The main purpose of implementing this kind of technology in to a store is to eliminate lines, since the products can be paid through the app without walking through a checkout. However, since the app can track the customer's movements and even what products the customer takes from the selves (and what she/he returns to the shelves after looking them), the technology can be utilized to bring the possibilities of bricks-and-mortar stores closer to online stores.

Another prospect for the future for bricks-and-mortar stores is to take advantage of customers' smartphones or other mobile technologies. For example, by using mobile phones customers could scan the products they are buying already while shopping. This would enable mobile checkouts but also enable the retailer to deliver product recommendations, coupons and promotions to the mobile phone while the customer is still shopping, and this way influence the customers purchasing decisions. (Aloysius, Hoehle, & Venkatesh, 2016)

For the future, there will be a lot of opportunities for further research in this field. The development of technology is very rapid and new technologies that brings more options for retailer are created all the time. For example, as mentioned in the previous section, gathering location based data about customers brings a whole lot new possibilities.

References

Adomavicius, G., & Tuzhilin, A. (2005). Personalization technologies: A process-oriented perspective. *Communications of the ACM*, *48*(10), 83–90.

Aguinis, H., Forcum, L. E., & Joo, H. (2013). Using Market Basket Analysis in Management Research. *Journal of Management*, *39*(7), 1799–1824.

Ahn, K. Il. (2012). Effective product assignment based on association rule mining in retail. *Expert Systems with Applications*, *39*(16), 12551–12556.

Aloysius, J. A., Hoehle, H., & Venkatesh, V. (2016). Exploiting big data for customer and retailer benefits A study of emerging mobile checkout scenarios. *International Journal of Operations & Production Management*, *36*(4), 467–486.

Arazy, O., & Kumar, N. (2010). A Theory-Driven Design Framework for Social Recommender Systems A Theory-Driven Design Framework for Social Recommender Systems. *Information Systems Journal*, *11*(9), 455–490.

Balanovic, Marko; Shoham, Y., Balabanović, M., & Shoham, Y. (1997). Content-Based, Collaborative Recommendation. *Communications of the Acm*, *40*(3), 66–72.

Bose, I., & Chen, X. (2009). Quantitative models for direct marketing: A review from systems perspective. *European Journal of Operational Research*, *195*(1), 1–16.

Chen, Y.-L., Tang, K., Shen, R.-J., & Hu, Y.-H. (2005). Market basket analysis in a multiple store environment. *Decision Support Systems*, *40*(2), 339–354.

Emrah, B., Kantardzic, M., & Cakir, O. (2015). Retail Store Segmentation for Target Marketing. In *Advances in Data Mining: Applications and Theorethical Aspects, ICDM 2015* (Vol. 9165, pp. 32–44). Hamburg, GERMANY.

Hu, Y. H., Huang, T. C. K., & Kao, Y. H. (2013). Knowledge discovery of weighted RFM sequential patterns from customer sequence databases. *Journal of Systems and Software*, *86*(3), 779–

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data Clustering: A review. *ACM Computing Surveys*, *31*(3), 264–323.

Kaur, M., & Kang, S. (2016). Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining. *Procedia Computer Science*, *85*, 78–85.

Keller, T., Raffelsieper, M., & Ag, C. (2014). Cosibon : An e-Commerce like platform enabling bricks-and-mortar stores to use sophisticated product recommender systems. In *8th ACM Conference on Recommender Systems* (pp. 367–368). Silicon Valley, Foster City: Association for Computing Machinery, Inc.

Kim, H. K., Kim, J. K., & Ryu, Y. U. (2009). Personalized recommendation over a customer network for ubiquitous shopping. *IEEE Transactions on Services Computing*, *2*(2), 140–151.

Kesko, Contractual terms and conditions. (n.d.). Retrieved May 02, 2017, from https://www.plussa.com/K-Plussa- in-English/#CONTRACTUAL- TERMS-AND- CONDITIONS

Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014). *Mining of Massive Datasets* (Second Edi). Cambridge : Cambridge University Press 2014.

Li, S. S., & Karahanna, E. (2015). Online Recommendation Systems in a B2C E-Commerce Context : A Review and Future Directions. *Journal of the Association for Information Systems*, *16*(2), 72–107.

Linoff, G. S., & Berry, M. J. (2011). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* (3rd ed.). John Wiley & Sons, Incorporated.

LVN, R., Wang, Q., & Raj, J. D. (2014). Recommending News Articles using Cosine Similarity Function. In *SAS Institute Inc. 2014. Proceedings of the SAS® Global Forum 2014 Conference* (pp. 1–7). SAS Institute Inc.

Miguéis, V. L., Camanho, A. S., & Cunha, J. F. (2011). Mining Customer Loyalty Card Programs : The Improvement of Service Levels Enabled by Innovative Segmentation and Promotions Design. In *Lecture Notes in Business Information Processing* (pp. 83–97). Springer Verlag.

Mitra, A., Ghosh, S., Basuchowdhuri, P., Shekhawat, M. K., & Saha, S. K. (2016). Recommendation system based on product purchase analysis. *Innovations in Systems and Software Engineering*, *12*(3), 177–192.

Nanda, K., & Ram, R. (2006). Using Basket Composition Data for Intelligent Supermarket Pricing. *Marketing Science*, *25*(2), 188–199.

Olson, D. L., & Chae, B. (2012). Direct marketing decision support through predictive customer response modeling. *Decision Support Systems*, *54*(1), 443–451.

Park, D. H., Kim, H. K., Choi, I. Y., & Kim, J. K. (2012). A literature review and classification of recommender systems research. *Expert Systems with Applications*, *39*(11), 10059–10072.

Raijas, A., & Järvelä, K. (2015). Päivittäistavaroiden ostaminen ja kuluttajien näkemykset kaupan kanta-asiakas järjestelmistä. Kilpailu- Ja Kuluttajaviraston Selvityksiä, 1

Päivittäistavarakauppa ry, Päivittäistavarakaupan tilastot. (n.d.). Retrieved May 02, 2017, from http://www.pty.fi/julkaisut/tilastot/

SOK, Asiakasomistaja- ja asiakasrekisteri. (n.d.). Retrieved May 02, 2017, from https://www.s-kanava.fi/web/s/s-kanavan- rekisteriseloste

Solnet, D., Boztug, Y., & Dolnicar, S. (2016). An untapped gold mine? Exploring the potential of market basket analysis to grow hotel revenue. *International Journal of Hospitality Management*, *56*, 119–125.

Videla-Cavieres, I. F., & Ríos, S. A. (2014). Extending market basket analysis with graph mining techniques: A real case. *Expert Systems with Applications*, *41*(4 PART 2), 1928–1936.

Vucetic, S., & Obradovic, Z. (2005). Collaborative filtering using a regression-based approach. *Knowledge and Information Systems*, *7*(1), 1–22.

Wei, J. (2010). Research on data preprocessing in supermarket customers data mining. In *2nd International Conference on Information Engineering and Computer Science - Proceedings, ICIECS 2010*.